What is Claimed:

1.   A clustering tree data structure encoded on a computer readable medium, comprising:

data that defines decision nodes associated with questions that relate to a group of sounds; and

data that defines terminal nodes that define a sound cluster to which the group of sounds belong;

wherein the decision nodes and the terminal nodes are defined hierarchically relative to one another and the decision nodes and the terminal nodes are divided into levels, each of the levels being associated with a different speech recognition model.

2.   The data structure of claim 1, wherein the sounds are represented by phonemes.

3.   The data structure of claim 1, wherein the group of sounds includes a sound being modeled and two context sounds before and after the sound being modeled.

4.   The data structure of claim 1, wherein decision nodes corresponding to lower ones of the levels in the hierarchically defined nodes are associated with more detailed questions than decision nodes corresponding to higher ones of the levels in the hierarchically defined nodes.

5.   The data structure of claim 1, wherein the decision nodes and terminal nodes are divided into three levels, including:

a first level associated with first questions that drive a triphone non-crossword speech model,

a second level associated with second questions that drive a quinphone non-crossword speech model, and

a third level associated with third questions that drive a quinphone non-crossword speech model.

6.   The data structure of claim 1, wherein there are significantly fewer sound clusters than possible sound groups.

7.   A method of building a clustering tree for classifying a group of sounds into one of a number of possible sound clusters, the method comprising:

building a first level of the clustering tree with a first hierarchical arrangement of decision nodes in which each of the decision nodes of the first hierarchical arrangement is associated with one of a first group of questions relating to the group of sounds; and

building a second level of the clustering tree with a second hierarchical arrangement of decision nodes in which each of the decision nodes of the second hierarchical arrangement is associated with one of a second group of questions relating to the group of sounds, the second group of questions

discriminating at a finer level of granularity within the group of sounds than the first group of questions.

8. The method of claim 7, further comprising:

freezing building of the first level of the clustering tree before building the second level of the clustering tree.

9. The method of claim 8, further comprising:

freezing building of the first level of the clustering tree when an entropy level of the first level of the clustering tree is below a predetermined threshold.

10. The method of claim 7, further comprising:

building a third level of the clustering tree with a third hierarchical arrangement of decision nodes in which each of the decision nodes of the third hierarchical arrangement is associated with one of a third group of questions relating to the group of sounds, the third group of questions discriminating at a finer level of granularity within the group of sounds than the second group of questions.

11. The method of claim 10, further comprising:

freezing building of the second level of the clustering tree before building the third level of the clustering tree.

12.   The method of claim 11, further comprising:

freezing building of the second level of the clustering tree when an entropy level of the second level of the clustering tree is below a predetermined threshold.

13.   The method of claim 7, wherein the clustering tree is further built to include terminal nodes that assign each of the groups of sound into one of the sound clusters.

14.   The method of claim 7, wherein the sounds are represented by phonemes.

15.   The method of claim 7, wherein the first group of questions includes questions that relate to the group of sounds as a sound being modeled and one context sound before and after the sound being modeled.

16.   The method of claim 15, wherein the second group of questions includes questions that relate to the group of sounds as the sound being modeled and two context sounds before and after the sound being modeled.

17. A speech recognition system comprising:

a clustering tree configured to classify a series of sounds into predefined clusters based on one of the sounds and on a predetermined number of neighboring sounds that surround the one of the sounds; and

a plurality of speech recognition models trained to recognize speech based on the predefined clusters, each of the plurality of speech recognition models receiving the predefined clusters from a different portion of the clustering tree.

18. The system of claim 17, wherein the different portions of the clustering tree correspond to hierarchical levels in the clustering tree.

19. The system of claim 18, wherein higher ones of the hierarchical levels include nodes that correspond to more general questions than questions corresponding to nodes at lower ones of the hierarchical levels.

20. The system of claim 17, wherein the plurality of speech recognition models include:

a triphone non-crossword speech recognition model;

a quinphone non-crossword speech recognition model; and

a quinphone crossword speech recognition model.

21.   The system of claim 17, wherein the sounds are represented by phonemes.

22.   The system of claim 17, wherein the series of sounds include a sound being modeled and two context sounds before and after the sound being modeled.

23.   The system of claim 17, wherein the clustering tree comprises:

decision nodes associated with questions that relate to the series of sounds, and

terminal nodes that define a sound cluster to which the series of sounds belong.

24.   The system of claim 23, wherein the decision nodes and the terminal nodes are defined hierarchically relative to one another and the decision nodes and the terminal nodes are divided into levels, each of the levels being associated with a different one of the plurality of speech recognition models.

25.   The system of claim 24, wherein the decision nodes correspond to lower ones of the levels in the hierarchically defined nodes are associated with more detailed questions than decision nodes corresponding to higher ones of the levels in the hierarchically defined nodes.

26.   A device comprising:

means for classifying a series of sounds into predefined clusters based on one of the sounds and a predetermined number of neighboring sounds that surround the one of the sounds; and

means for training a plurality of speech recognition models to recognize speech based on the predefined clusters, each of the plurality of speech recognition models receiving the predefined clusters from the means for classifying.